守正能否创新？
——基于我国图情领域论文新颖性和传统性的分析

■ 梁兴堃
北京大学信息管理系 北京 100871

摘 要：[目的/意义]以我国图情领域为例，测量论文的新颖性和传统性并探究其对论文学术影响力的作用进而揭示学术创新的规律。[方法/过程]采用基于马尔科夫链蒙特卡罗(Markov chain Monte Carlo, MCMC)的方法，对我国2000年至2019年20年间在中文社会科学引文索引(CSSCI)中收录的图书馆学情报学领域的70 207篇研究论文的新颖性、传统性进行测量，并分析论文新颖性和传统性对论文学科影响力的作用。[结果/结论]结果显示，其他因素不变时，论文新颖性提高1个单位，论文成为高被引论文的优势比增加11%，而论文传统性提高1个单位，论文成为高被引论文的优势比增加33%。边际效应分析显示，同时具有较高的新颖性和传统性的论文较之于其他类型的论文具有更高的成为高被引论文的可能性。此外，随着时间推移，新颖性对论文成为高被引论文概率的影响逐渐削弱，而传统性的影响逐渐增强。同时，作者团队规模对于论文的新颖性存在显著影响，这种影响随着时间的推移而增强。这些发现凸显我国图情领域守正创新的特点，为理解我国图情领域的学术创新规律提供新的实证基础。同时，也提出一种不同于传统信息计量的基于贝叶斯统计的新方法。

关键词：新颖性 传统性 学科影响力 高被引论文 论文被引量 马尔科夫链蒙特卡洛方法(MCMC) 团队作者 贝叶斯统计
分类号：G25

1 研究背景

长期以来有关图情领域创新发展的研究聚焦于宏观方向的把握和梳理[1-8]，而忽视了从微观上探究图情领域科研人员的学术创新机制。从微观机制来看，通常认为，学术创新是科研人员创造性地利用已有知识而创造新知识或新想法的过程[9-13]。科研人员可以选择创造新颖知识以使新的研究成果能突破已有研究成果，而使其具有新颖性[9,11, 14-15]。同时，科研人员在学术创新的过程中也会选择创造性地重组已有知识，从而使得新的研究成果更好继承、整合已有的研究成果，使其更具传统性[9-10,12,14, 16-17]。国外的研究表明，一项具体的研究成果通常兼具新颖性和传统性，但二者的程度可能不尽相同；而不同程度的新颖性与传统性的组合会使得学术成果产生不同的学科影响力[9,14-16]。因此，探究图情领域学术创新规律的一个核心问题是图情领域的科研人员如何利用兼具新颖性和传统性的知识产生有影响力的创新成果。

虽然这一问题对于理解图情领域的创新发展非常重要，但一直以来缺乏系统的实证研究。而国外的相关研究尽管可以借鉴，但不同文化背景、学科领域下，学术创新的微观机制有着明显差异[14-15,18]。因此，需要系统深入研究这一问题才有助于全面系统地把握图情领域创新发展的规律。为此，笔者以2000年至2019年20年间在中文社会科学引文索引(CSSCI)中收录的我国图书馆学情报学领域的70 207篇研究论文为研究对象，采用基于马尔科夫链蒙特卡罗(MCMC)的方法测量这些论文的新颖性和传统性，并分析论文新颖性和传统性对论文学科影响力的作用及其随时间的变化趋势。此外，笔者还进一步从作者团队的角度探究论文的新颖性和传统性形成的微观机制，为全面理解图情领域学术创新的微观机制提供实证支持。

作者简介：梁兴堃，助理教授，博士，E-mail：lxk@pku. edu. cn。
收稿日期：2022-08-01 修回日期：2022-09-21 本文起止页码：148-161 本文责任编辑：易飞

2 研究假设

2.1 研究成果新颖性、传统性与其学术影响力的关系

新颖性是学术成果的重要属性。通常认为,新颖性是相对的,是相较于一定标准而言更为罕见的内容[9,14-15,18]。这里的"一定标准",可以是某一学科已有的知识[9-10],也可以是一个时代的知识体系[18]。例如,用先进的算法进行数据分析和挖掘在图情领域已经不具有新颖性,但在其他一些学科中可能依然非常新颖。新颖性强的研究成果往往需要研究者通过其创造力产生具有创新性甚至突破性的新思想或新见解,如新理论、新研究方法、新研究主题等[9,14,17]。这些新思想或新见解往往能启发相关领域的研究者,进而产生更强的学科影响力[9,11]。B. Uzzi 等对 Web of Science 中来自不同领域的 1 790 万篇论文的分析发现,在控制一些相关因素后,新颖性高的论文成为高被引论文的概率明显要高[9]。后续针对其他类型的英文研究成果的新颖性开展的研究均得出基本一致的结论[14-18]。

新颖性的这一作用在我国图情领域的研究中也有所体现。例如,我国图书馆学基础理论发展的早期,不少有影响力的理论都是本土化理论,是结合我国实际发展和扬弃西方图书馆学理论而形成,具有明显的新颖性[19];在信息技术发展变革时期,数字图书馆、智慧图书馆等新概念、新理念出现后就得到学界大量的关注,相继成为研究前沿[20],而与之呼应的新出现的知识服务、智慧服务等也成为研究热点,在学界产生了广泛的影响[21]。据此,得出如下假设:

假设1:图情领域的研究论文,新颖性越强,学科影响力越大。

尽管新颖性体现了科研中的创造性,但不少研究指出,对于科研人员而言,新颖性会是一把双刃剑[22]。一方面,追求新颖性会增加学术研究的不确定性,可能会导致研究者生产力的降低和研究周期的增长[17,23];另一方面,受同行交流等因素影响,新颖性强的研究成果在短时间内更难被理解和认可,从而产生一系列负面效应[15-25],如更低的论文发表率[17]、研究成果的学科价值和贡献被低估并导致这些论文的作者被逐渐边缘化[24]。此外,新颖性高的研究计划在科研基金的评审中也更容易被误解而得到低分,被淘汰出局[25]。对新颖性的这种"偏见",并非是科研领域独有的,在其他领域也存在。例如,在商业领域也存在个人或集体对新颖性观点的偏见[26-28]。一些研究指出,这种偏见产生的根源除了商业领域追求降低成本和实用性而形成的有限理性,更在于人们对于新观点的理解机制——人们在理解新观点时,会利用自身的经历、知识等来弥补新观点与已有认知之间的信息缺口(informational gap),而新颖性强的内容造成的信息缺口不容易在这个机制下得到弥补,从而导致其被理解和接受的困难[28]。

由于新颖性是一把双刃剑,研究者往往在追求新颖性的同时,保持一定的传统性。所谓"传统性",并非是新颖性的对立面,而表现为与已有知识和理论体系的紧密关联。传统性的增加有助于弥补新颖性的内容造成的信息缺口,从而使新颖性的内容更容易地被接受[15,28]。从另一个角度来看,科学研究中新颖内容的产生离不开对已有知识或理论的传承和发展,即通过对已有知识、理论、技术或方法的梳理归纳,进而形成具有创造性的内容。这一特点在图情领域的研究中亦有体现。例如,杜威在论述图书馆员职业教育的必要性时,并没有从其新创立的十进制分类法的角度展开论述,而是以图书馆员服务读者、图书馆社会教育等业内业外更熟悉的角度出发[29];王重民在阐述《四库全书总目》的目录学价值时,并没有直接介绍《四库全书总目》的内容特点,而是从大家更为熟悉的《四库全书》编纂历史背景出发,在介绍《四库全书总目》编纂过程的基础上,分析其目录编纂方法的特点[30]。总之,图情领域学术成果的创新离不开对已有知识的传承和发展[31]。图情领域的研究者在追求新颖性的同时,也需要在传统性上下功夫,以体现学术传承和积淀[9,15]。因此,传统性突出的研究成果有助于帮助其他研究者理解现有研究的发展脉络,也能产生更广泛的学科影响力。由此,得出如下假设:

假设2:图情领域的研究论文,传统性越强,学科影响力越大。

2.2 团队作者与研究成果的新颖性、传统性的关系

依时间维度横向对比可以发现,论文的篇均作者数量在不断增加;同时,随着科研复杂度的不断提高,当前高影响力的研究成果几乎都是以团队的形式完成[9]。对团队作者的研究有助于厘清研究成果新颖性和传统性形成的微观机制。不少研究认为,作者团队规模的扩大有助于形成更具多样性的团队,从而将不同学科领域、不同深度和广度的知识融合,进而有助于突破固有的研究边界,产生更具创新性和影响力的学术成果[9-14,32-33];同时,合作者团队成员之间的互相选择机制有助于保障研究成果质量的下限,即减小产生

很差的学术成果的可能性[34]。此外,作者团队规模扩大也带来了团队协调的成本增加和沟通困难等负面影响[32-33,35],因此,这些研究普遍认为保持合理的团队规模更有助于创新。我国图情领域科研合作的特点与这些研究基本一致,呈现出论文以 2-3 人合作为主的趋势,并且越来越多的高影响力论文是合作作者完成的[36-38]。因此,可以得出假设:

假设 3:在图情领域,团队作者的研究论文更具学科影响力。

3 方法与数据

3.1 数据来源及其预处理

中文论文数据库有多种来源,笔者选择 CSSCI 数据库,因为该数据库的数据质量较之于万方、维普、知网等商业数据库更好,特别是该数据库将作者和作者单位进行了一一对应,降低了消歧的难度。笔者选择以 CSSCI 数据库收录的所有图书馆学情报学领域的期刊在 2000 年至 2019 年 20 年间所发表的所有论文作为样本,并对其进行数据清洗。首先,去掉样本中的通知、社论、勘误等研究价值不大的文献类型,最终保留 70 207 篇研究论文或综述。然后,对样本中的作者和作者单位进行消歧。作者消歧主要利用部分论文的作者简介中的出生日期信息,并结合论文自引和是否存在共同合作者的情况综合判断。作者单位信息并非本文关注的重点,因此只在研究机构层面(校级),利用教育部公布的机构名称变动公告对机构名称变更进行处理,确保机构名称的统一。最后,通过抽查的方式进行人工核验确保作者和作者单位的消歧质量。

3.2 变量及其测量

3.2.1 因变量:学科影响力

通常而言,不少研究以论文被引频率或百分位数来测量论文的学科影响力,越来越多的研究强调百分位数的优势[39-42],但学界还没有形成广泛认可的标准来界定被引量在前百分之几的论文为高被引论文,或者说具有广泛影响力的论文。在研究实践中,通常选择一个学科中论文被引量的前 1%、前 5% 或前 10% 等不同的标准来界定[9-14,17-18,24]。参照相关研究[9-14],并考虑到 CSSCI 数据库中的被引关系更为稀疏,笔者选择被引量前 5% 作为高被引论文的标准来测量论文的学科影响力。同时,以被引量前 1% 和前 10% 作为替代测量以检验该标准的统计稳健性。此外,由于图情领域引用关系的稀疏性,高被引论文通常只能说明非常小的一部分论文的学科影响力,而不能说明一般论文的影响力[39,42]。为了更稳健地测量学科影响力,笔者还选择论文被引总量作为学科影响力的替代测量进一步检验模型的统计稳健性。

3.2.2 自变量:新颖性和传统性

笔者采用一种优化后的 MCMC 方法以测量新颖性和传统性[9,14]。这种方法认为,新颖性或传统性都是一个相对的概念,不能直接以观测到的频率值作为新颖性或传统性的测量,而需与一定的标准进行比较;一个合适且相对客观的方式是与随机概率进行比较[9,14-15,18],即排除观测值中"常见"或"不常见"的偶然因素[9-14,17],进而说明内容的新颖性。例如,一个概念在一个学科中出现 10 次,究竟是"新",还是"不新",无法直接从可观测的"出现 10 次"中得出结论,而需要与将"10 次"与一个随机产生的"新概念"在这个学科中可能出现的次数进行比较来说明其新颖性。

以此为出发点,这些研究认为论文的参考文献很好地体现了论文的学科基础及其对已有知识的利用情况,科学研究成果的新颖性(或传统性)体现在论文的共引文献的新颖程度(或传统程度)上。通过测度共引文献的新颖程度(或传统程度)可以得到科学研究成果的新颖性(或传统性)的程度。具体算法是:

(1)根据论文的共引文献整理其中所有出现的期刊的两两组合。

(2)整理所有论文共引文献中的期刊两两组合,按年度记录下各组合的频率 o。

(3)根据论文间的引用关系,构造引文网络。与此同时,为了得到可以比较的"随机概率",这些研究采用基于 MCMC 的方法,通过将观测到的引文网络随机化,来模拟出可比较的引文网络。由于 MCMC 的方法对算力要求较高,B. Uzzi 等采用吉布斯采样法(Gibbs sampling)的方法来随机化观测到的引文网络[9],并建议通过交换观测到的引文网络中各个子图内不少于 $n_i \times 2\log_{10} n_i$ 次的引用关系,其中,n_i 为该引文网络中第 i 个子图的引用关系总数,使 MCMC 收敛或近似收敛以得到随机化后的引文网络。该网络的节点数、边数、各个节点的入度和出度均与观测到的网络一致,且随机化后被引论文被引的年份及其引用来源论文发表的年份都没有变化。这里需要注意两点:①吉布斯采样方法并不是严格随机;②B. Uzzi 等建议的交换次数并非是 MCMC 达到收敛的必要条件,而是在对算力要求较高时可以采取的一个经验法则。

(4)将步骤(3)重复 10 次,得到 10 个随机化后的引文网络。统计每个随机化后引用网络中每篇论文的

引用关系,得出各期刊组合被共引的频率,计算该频率的均值 e 与标准差 σ,然后计算各期刊两两组合的标准化分数(z),计算公式如下:

$$z = \frac{o - e}{\sigma} \qquad \text{公式(1)}$$

(5)根据某一论文共引期刊两两组合的标准化分数 z 的分布,选择合适的百分位数作为该论文新颖性和传统性的测量。该方法的创新性在于:现实中可观测的频率值只有一个,而通过 MCMC 的方法可以将观测到的引文网络随机化,并通过计算标准化分数(z 分数)以消除随机因素带来的影响。由于均值 e 是基于随机概率得到的,因此,得到的 z 分数小于 0 的组合是相较于随机概率而言更为罕见的组合,具有一定的新颖性,而标准化分数大于 0 的组合,则是更为"传统"的组合。图 1 展示了任选的一篇论文(论文为吴慰慈与谷秀洁于 2007 年发表于《图书馆论坛》的《2006 年图书馆学学术进展》)共引的期刊两两组合的观测值和模拟值的累计分布情况(此处模拟值的计算采用的是在后文中阐释的优化后的方法,且模拟值为 10 次模拟后得到的均值)。由图 1 可知,观测值较为稀疏,而模拟值分布相对密集;模拟值类似于随机产生的"基准",与这个基准相比较可以更好地理解观测值的相对新颖程度。

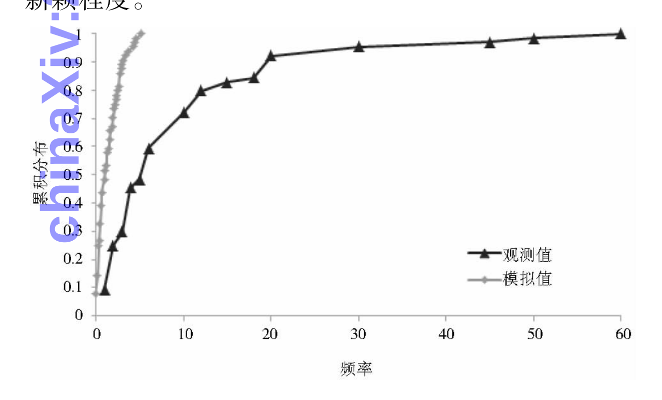


图 1 一篇论文中可观测到的论文共引期刊两两组合出现的频率与通过随机模拟得到的论文共引期刊两两组合出现的期望频率

笔者对新颖性和传统性的测量进行了优化。不少研究强调,引文网络具有很强的偏好依附(preferential attachment)特性[43-46],即被引用过的论文更容易获得更高的被引量。因此,在将实际观测到的引文网络随机化的过程中应当保持引文网络的这种特性,而不能假定论文的引用(或不被引用)是一个随机过程。为了更好地模拟这个特性,笔者借鉴随机过程中基于狄利克雷过程的中餐馆过程(Chinese restaurant process)[47],具体算法是:

(1)初始情况下,假定每篇论文被引的概率是均等的。

(2)采用吉布斯采样法,选择任意的两组观测到的存在引用关系的论文,在交换其引用关系时,先考虑论文被引的概率 p 是否大于一个 0 到 1 之间的随机值 u,如果大于 u 则交换引用关系,如果小于 u 则不交换引用关系。

(3)在后续过程中,利用狄利克雷分布的概率密度函数来修正随机值 u,即根据观测到的论文和期刊被引的情况调整论文和期刊的引用偏好,使得某一论文被引概率会随该论文被引次数和论文所在期刊被引次数的增加而增加。

(4)重复步骤(2)和(3),直至所有引用关系都至少交换一次。这里需要注意,由于本研究的引文网络规模中等,因此可以采用全部交换的方式。如果引文网络规模太大,交换所有引用关系的算力成本会非常大,则此时可以考虑使用 B. Uzzi 等推荐的经验法则。

为了直观地说明新颖性和传统性的测量,因此随机选取了一篇论文(论文为马费成和裴雷于 2005 年发表在《情报学报》的《我国信息资源共享实践及理论研究进展》)并在图 2 中展示该论文参考文献的共引期刊组合的标准化分数的分布情况。从图 2 可看出,在前 50 百分位数之前,标准化分数的整个分布较为长尾,而在其之后,分布更为密集。这说明相对新颖的期刊组合在文献中出现的概率的确比较稀疏,而相对传统的期刊组合出现的概率比较频繁。这也从侧面说明标准化分数的第 10 百分位数的期刊组合确实出现得"更罕见",新颖性更强。

为说明新颖性和传统性的整体情况,将 2000 年至 2019 年分成两个 10 年(2000s 和 2010s),分别统计了这两个 10 年期间发表的论文的新颖性(其参考文献中所有期刊组合的标准化分数的第 10 百分位数,见图 3)和传统性(其参考文献中所有期刊组合的标准化分数的第 50 百分位数,见图 4)的情况。从图 3 可以看出,在 2000-2019 年间,我国图情领域论文整体的新颖性在提高,其累计分布比较稀疏,第 10 百分位数对应的期刊组合也确实更为"新颖"。在 2000-2009 年间(n=28 890),只有 14.76% 的论文其新颖性是小于 0 的,而在 2010-2019 年间(n=41 317),有 28.83% 的论文其新颖性小于 0,增长明显。从图 4 可以看出,在 2000-2019 年间,我国图情领域论文整体的传统性保持在较高的水平,其累计分布在标准化分数较小的区

间更为稀疏,呈现长尾的特点。在 2000－2009 年间,图情领域的论文中只有 2.23% 的论文的传统性小于0,而在 2010－2019 年间,这个比例变化不明显,为 2.32%。这些结果说明,标准化分数分布中的第 10 百分位数可以较好地反映其新颖性,而第 50 百分位数可以较好地反映其传统性。此外,这些结果与 B. Uzzi 等的研究结果基本类似[9],但相较之下,新颖性的比例要明显低得多,而传统性的比例明显更高。

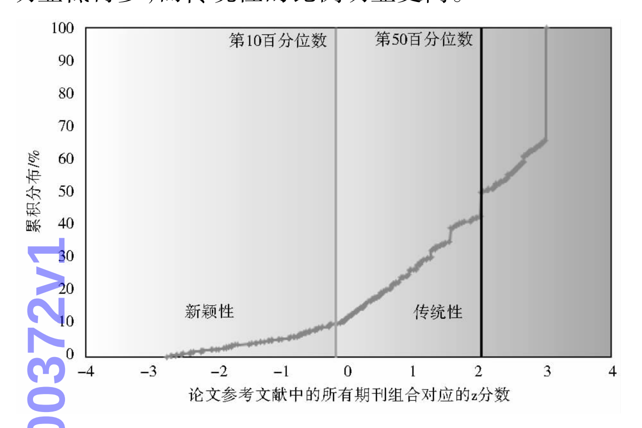


图 2 论文引用的期刊两两组合的标准化分数分布及其与新颖性和传统性关系图例

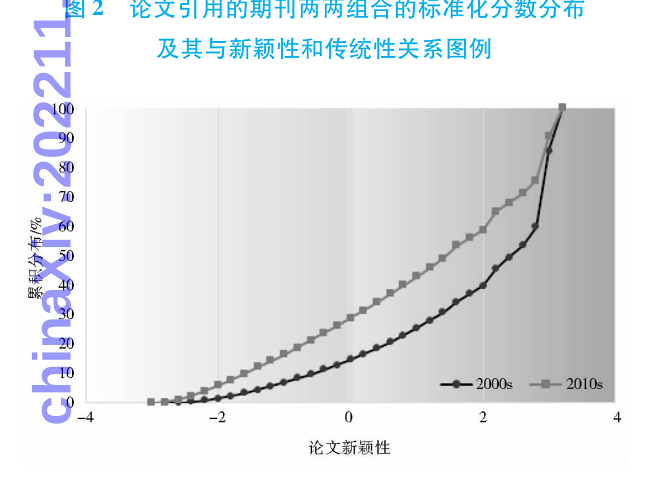


图 3 2000－2009 年和 2010－2019 年两个 10 年间论文新颖性的整体情况

3.2.3 控制变量

参照以往研究,为了排除潜在的替代解释,在回归模型中加入论文发表期刊、论文发表年份和论文作者数量 3 个变量作为控制变量。其中,论文发表期刊是分类变量,论文发表年份是自然年份的分类变量,在模型中控制其固定效应。参照 B. Uzzi 等提出的团队作者规模的分类方法以及我国图情领域合作者的特点[9,36],将论文作者数量分为独立作者、双人合作和团队合作(3 人及以上)3 类。

3.3 回归模型及分析策略

与研究假设对应,构造多元回归模型,公式如下所示:

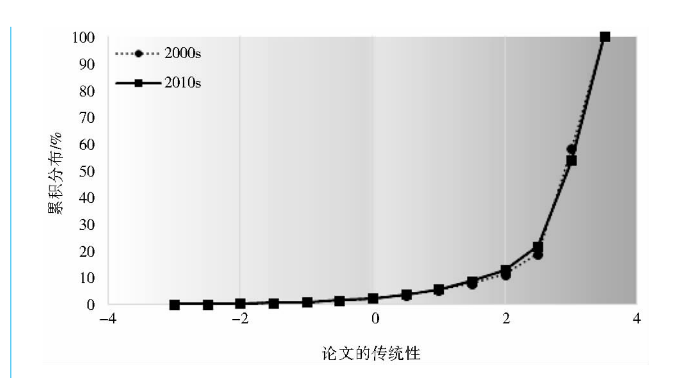


图 4 2000－2009 年和 2010－2019 年两个 10 年间论文传统性的整体情况

$$学科影响力_t = \beta_1 \times 新颖性_t + \beta_2 \times 传统性_t + \beta_3 \times 作者数量_t + \beta_4 论文发表期刊 + \beta_0 + \varepsilon \quad 公式(2)$$

其中,β_0 为常数项,ε 为随机误差,下标 t 代表论文发表年份,β_1 到 β_4 为待估计系数,且,$\beta_1 > 0$,$\beta_2 > 0$,$\beta_3 > 0$。

利用上述模型以 logistic 回归检验论文新颖性、传统性和作者数量对论文成为高被引论文概率的影响,从而说明这些因素与论文学科影响力的关系。需要说明的是,使用标准化分数测量新颖性时,标准化分数越小,新颖性越强。因此,如果要检验新颖性越强则论文成为高被引论文的概率越高这一假设,而将标准化分数累积分布的前 10 百分位数直接作为新颖性的测量标准进行回归分析时,会因为人为因素导致 β_1 为负。这容易造成对回归分析结果的误解。鉴于此,在回归分析时,选择标准化分数累积分布前 10 百分位数的相反数作为新颖性的测量标准,这样就能保证该变量值越大,新颖性越强。此外,在分析过程中,严格控制出版年和作者数量的固定效应。logistic 回归分析结果见表 1。

4 分析结果

4.1 新颖性、传统性与论文学科影响力的关系

表 1 显示了以论文被引量前 5% 为标准来界定高被引论文的 logistic 回归分析结果。其中,模型 1 使用全部论文样本,模型 2 只使用 2000－2009 年间的论文样本,模型 3 只使用 2010－2019 年间的论文样本。整体来看,论文的新颖性对其成为高被引论文优势比(odds ratio)的提升非常显著($\beta = 0.109, p < 0.0001$),论文的传统性对其成为高被引论文优势比的提升也非常显著($\beta = 0.292, p < 0.0001$)。对比 2000－2009 年和 2010－2019 年的两个 10 年间的情况可以看出, 论

表 1 logistic 回归分析结果

因变量	模型 1 高被引论文(前 5%)	模型 2 高被引论文(前 5%)	模型 3 高被引论文(前 5%)
新颖性	0.109***	0.120***	0.102***
	(0.013)	(0.022)	(0.016)
传统性	0.292***	0.241***	0.347***
	(0.035)	(0.052)	(0.047)
双人作者	0.359***	0.246***	0.474***
	(0.041)	(0.059)	(0.059)
团队作者	0.472***	0.373***	0.569***
	(0.047)	(0.076)	(0.062)
发表期刊	-0.003***	-0.003***	-0.002***
	(0.000)	(0.000)	(0.000)
发表年份固定效应	是	是	是
常数项	-2.466***	-2.165***	-4.195***
	(0.127)	(0.157)	(0.165)
log likelihood	-13 065.182	-5 979.789	-7 073.442
Chi2	1 138.918***	479.137***	673.354***
样本量	70 207	28 890	41 317

注:①3 个模型中的系数均为初始系数;②括号内为稳健标准误;③显著性水平标记:* $p<0.05$, ** $p<0.01$, *** $p<0.001$

文新颖性对其成为高被引论文优势比的影响在减弱,而论文传统性对其成为高被引论文优势比在增强。这些结果都为说明新颖性和传统性对学科影响力的作用提供了有力的实证证据。

与线性回归不同,在 logistic 回归中,因变量与自变量的关系是非线性的,两个自变量对因变量的影响不等价于两个自变量系数的累加[48]。为进一步分析新颖性、传统性对学科影响力的作用,分别将新颖性(或传统性)的第 25 百分位数和第 75 百分数作为其高值和低值,可以形成新颖性高且传统性高、新颖性低且传统性高、新颖性低且传统性低、新颖性高且传统性低 4 种不同的组合情况,进而可以分析每种情况下新颖性和传统性对学科影响力的联合边际作用。分析结果如图 5 所示,当新颖性和传统性都很高时,论文成为高被引论文的概率是 7.08% ($p<0.000\ 1$);在新颖性和传统性都很低时,论文成为高被引论文的概率是 3.15% ($p<0.000\ 1$);新颖性高而传统性低时,论文成为高被引论文的概率是 4.90% ($p<0.000\ 1$);新颖性低而传统性高时,论文成为高被引论文的概率是 4.60% ($p<0.000\ 1$)。换言之,新颖性高的论文,无论其传统性高低都有更高的概率成为高被引论文。

4.2 论文作者数量与论文学科影响力的关系

从表 1 的模型 1 中可以看出,相较于独立作者的论文,双人作者的论文成为高被引论文的优势比要高

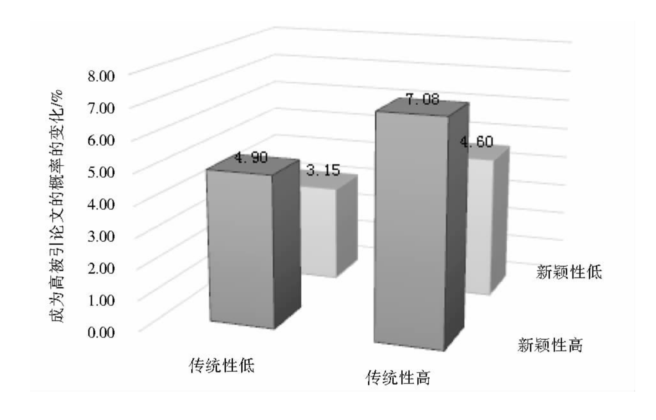


图 5 论文新颖性和传统性的不同组合对其成为高被引论文概率的影响

43.17% ($p<0.000\ 1$),团队作者的论文成为高被引论文的优势比独立作者论文的概率要高 60.33% ($p<0.000\ 1$)。对比表 1 的模型 2 和模型 3,可以进一步看出,在 2000 年至 2019 年,作者数量对于我国图情领域的论文成为高被引论文的影响在增强。这充分说明了团队作者在学科创新中的优势,并且随着时代的发展,这种优势在扩大。

为了进一步探究作者论文数量的影响,笔者统计了两个 10 年间独立作者、双人作者和团队作者的所有论文的标准化分数前 10、15、20、25、30、35、40、45 和 50 百分位数的均值,结果见图 6。从整体来看,随着作者数量的增加,论文的标准化分数的前 35 百分位数基本都呈现下降的趋势,而第 35 百分位数之后的变化不显著。如果对比前后两个 10 年,不难发现,随着时代的发展,无论是独立作者、双人作者还是团队作者,其整体的标准化分数都在下降。此外,在 2000-2009 的 10 年间,独立作者、双人作者和团队作者在这些标准化分数的均值上差距不大;而在 2010-2019 年的 10 年间,独立作者、双人作者和团队作者在这些标准化分数的均值差距逐渐扩大。这些变化趋势都在前 35 百分位数,特别是新颖性上,体现得更为明显。这可能意味着,对我国图情领域而言,作者数量的增加有可能会促进论文的新颖性,而对传统性的影响不大。

4.3 分析结果的稳健性

模型整体的共线性程度较低。在控制论文发表年份的固定效应后,表 1 中模型 1 的方差膨胀因子为 2.63,主要自变量的方差膨胀因子均小于 1.7,因此模型不存在多重共线性的问题。而异方差检验(怀特检验)的卡方值为 1 819.22($p<0.001$),这说明存在异方差问题,因此,在回归分析时,均使用稳健标准误代替一般的标准误[49]。

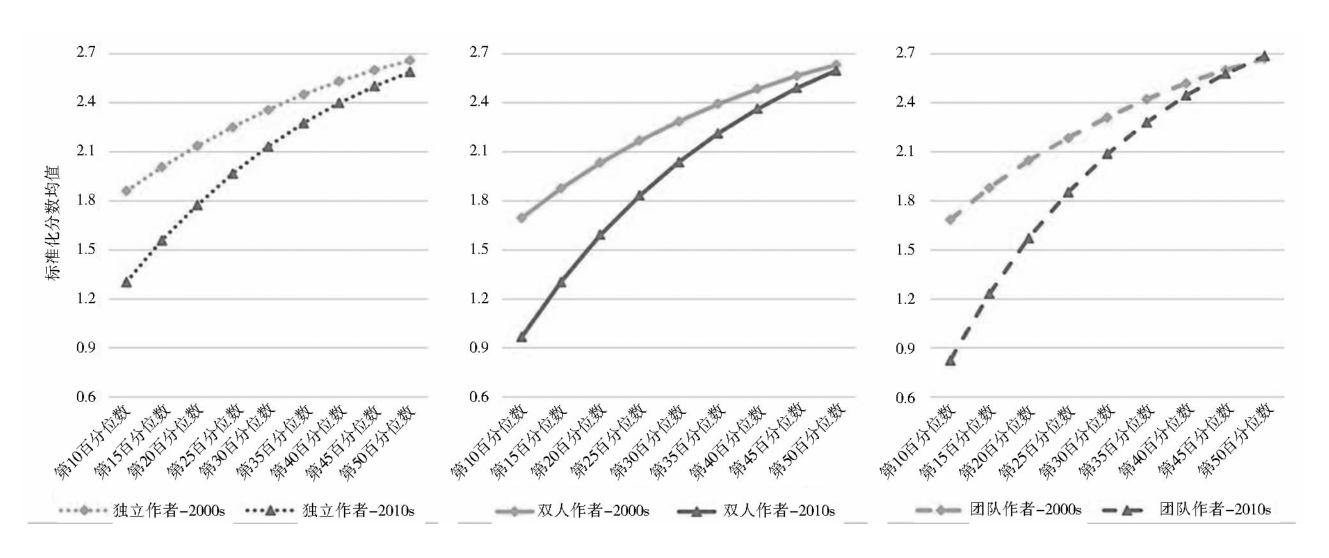


图 6　不同时期独立作者、双人作者和团队作者的作品的期刊两两组合分布的标准化分数前 50 百分位数的平均水平

此外，使用因变量和自变量的多个替代测量来检验回归分析结果的稳健性。首先，用论文发表后 4 年内的被引总量作为因变量，以相同的模型设定进行负二项回归（见表 2 的模型 1）。采用论文发表 4 年内的被引量，主要依据是我国图情领域文献半衰期大约为 3－4 年。得到的结果与基准模型基本一致，新颖性和传统性的系数正负及其显著性都没有变化。其次，将高被引论文的界定标准变为被引量前 1% 的论文和被引量前 10% 的论文，以相同的模型设定进行 logistic 回归（见表 2 模型 2、模型 3），结果也与基准模型基本一致，新颖性和传统性的系数正负及其显著性无明显变化，仅在模型 2 中，新颖性系数的显著性略有下降（p = 0.019）。

再次，选择标准化百分位数累计分布的前 15、20、25、30 百分位数作为新颖性的替代测量，回归分析结果见表 3。从表 3 可以看出，采用不同的标准时，新颖性对论文成为高被引论文的影响非常显著，各模型对应系数的 p 值均小于 0.001。这些结果充分说明了原分析结果的统计稳健性。

表 2　因变量选择不同标准测定的回归分析结果

因变量	模型 1 论文被引次数	模型 2 高被引论文 （前 1%）	模型 3 高被引论文 （前 10%）
新颖性	0.062 ***	0.066 *	0.102 ***
	（0.005）	（0.029）	（0.010）
传统性	0.141 ***	0.310 ***	0.246 ***
	（0.010）	（0.080）	（0.024）
双人作者	0.216 ***	0.529 ***	0.353 ***
	（0.016）	（0.089）	（0.030）
团队作者	0.295 ***	0.576 ***	0.501 ***
	（0.017）	（0.103）	（0.034）
发表期刊	－0.001 ***	－0.003 ***	－0.002 ***
	（0.000）	（0.000）	（0.000）
发表年份固定效应	是	是	是
常数项	0.744 ***	－1.781 ***	－4.025 ***
	（0.052）	（0.092）	（0.270）
lnalpha	0.437 ***	－	－
	（0.011）		
log likelihood	－116 041.64	－21 546.902	－3 677.703
Chi2	3 812.503 ***	1 752.522 ***	348.363 ***
样本量	70 207	70 207	70 207

注：①模型 1 为负二项回归，模型 2 和模型 3 为 logistic 回归，模型中的系数均为初始系数；②括号内为稳健标准误；③显著性水平标记：* $p<0.05$，** $p<0.01$，*** $p<0.001$

5　讨论

5.1　新颖性与传统性的测量

5.1.1　两种测量方法的比较

由于引文网络的特点，笔者采用一种优化后的基于 MCMC 的测量新颖性与传统性的方法。通过与 B. Uzzi 等提出的方法得到的结果进行比较，发现优化后的方法具有一定的优势。首先，计算各个年份的所有期刊两两组合的标准化分数，并比较两种方法得到的结果均值、标准差和最值等（见表 4）。从表 4 可以看出，两种方法的结果似乎没有明显差异。然后，进一步计算两种方法得到的结果的相关性，相关系数为 0.797 5（n = 24 332 554，p < 0.000 1）。

表 3　自变量选择不同标准测定的 logistic 回归分析结果

因变量	模型 1 高被引论文	模型 2 高被引论文	模型 3 高被引论文	模型 4 高被引论文
新颖性－第 15 百分位数	0.098 *** （0.015）			
新颖性－第 20 百分位数		0.080 *** （0.017）		
新颖性－第 25 百分位数			0.073 *** （0.021）	
新颖性－第 30 百分位数				0.069 ** （0.027）
传统性	0.279 *** （0.036）	0.255 *** （0.037）	0.243 *** （0.040）	0.236 *** （0.044）
论文作者数量(基线:独立作者)				
双人作者	0.365 *** （0.041）	0.371 *** （0.041）	0.373 *** （0.041）	0.375 *** （0.041）
团队作者	0.484 *** （0.046）	0.494 *** （0.046）	0.498 *** （0.046）	0.502 *** （0.046）
发表期刊	－0.003 *** （0.000）	－0.003 *** （0.000）	－0.003 *** （0.000）	－0.003 *** （0.000）
发表年份固定效应	是	是	是	是
常数项	－2.447 *** （0.127）	－2.417 *** （0.126）	－2.399 *** （0.126）	－2.385 *** （0.126）
log likelihood	－13 077.766	－13 087.899	－13 092.118	－13 094.338
Chi2	1 120.040 ***	1 103.697 ***	1 097.467 ***	1 095.282 ***
样本量	70 207	70 207	70 207	70 207

注:①各模型中的系数均为初始系数;②括号内为稳健标准误;③显著性水平标记: $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$

表 4　两种基于 MCMC 的方法将引文网络随机后标准化分数的比较

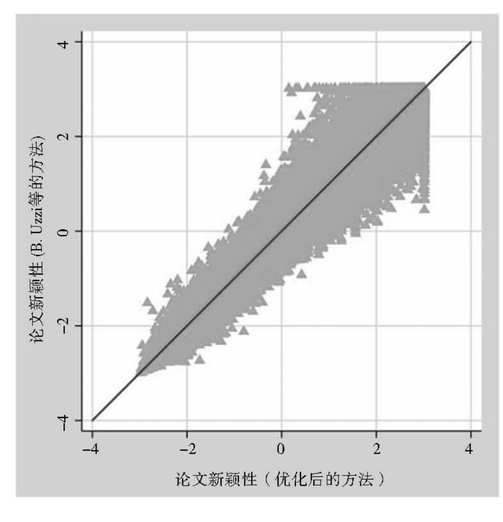
方法	各期刊组合标准化分数							
	描述性统计				两种方法结果的相关性分析			
	均值	标准差	最小值	最大值	全部样本	前 10% 样本	前 30% 样本	前 50% 样本
本文提出的方法	2.175	1.281	－3.015	3.015	0.797 5 ***	0.662 4 ***	0.703 1 ***	0.727 0 ***
B. Uzzi 等提出的方法	2.176	1.279	－3.011	3.015				

注:①显著性水平标记: $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$

为了更进一步研究这种差异性,将所有论文通过两种不同方法测得的新颖性和传统性以散点图的形式呈现(见图 7)。从图 7 可以看出两种方式的差异。通过配对样本 t 检验,发现两种方法得到的结果方差齐次,但优化后方法测出的新颖性(标准化分数的第 10 百分位数)均值显著低于 B. Uzzi 等提出方法得到的结果($T=2.22$, $p=0.013$,单边检验)。同时,优化后的方法得到的标准化分数的第 15、第 20 百分位数也均低于原方法得到的结果(T 值分别为 1.65,1.99,p 值分别为 0.049,0.023,单边检验),而第 25、第 30、第 35、第 40、第 45 和第 50 百分数的标准化的均值不存在显著性差异。这说明两种方法得到的结果的整体分布具有一致性,但对于标准化分数靠前的 20% 存在一定的差异。为了进一步说明这种差异,将两种方法的结果按不同的百分位数进行相关性分析。结果发现,两种方法得到的结果的标准化分数均排在前 10%,其相关性为 0.662 4($n=1\ 675\ 706$, $p<0.000\ 1$);均排在前 30% 的,其标准化分数的相关性为 0.703 1($n=5\ 212\ 511$, $p<0.000\ 1$);均排在前 50% 的,其标准化分数的相关性为 0.727 0($n=9\ 871\ 118$, $p<0.000\ 1$)。这进一步说明了两种方法得到的结果存在一定的差异性,并且在标准化分数靠前的部分差异更明显。这意味着优化后的方法对新颖性更敏感,更适应稀疏的引文网络。

5.1.2　新颖性、传统性测量的适用性

为了直观地说明新颖性或者传统性测量在我国图情领域的适用性,笔者选择 2000 年和 2019 年的一些代表性的期刊组合,按照标准化分数高低列在表 5 中。从表 5 可以看出,这些期刊组合出现在参考文献中的“偶然程度”(即新颖性)与预期基本一致。例如,《中国图书馆学报》和《大学图书馆学报》,在 2000 年和 2019 年经常同时作为参考文献出现。在 2000 年,《图书馆建设》和《情报理论与实践》的组合,相较于随机概率而言,是较为新颖的。图情领域与计算机科学的结合在 2000 年左右是具有一定新颖性的,在 2000 年《情报学报》与《人工智能学报》的组合的标准化分数恰好为 0,而在 2019 年,这种结合已经是“司空见惯”,如在 2019 年, *Journal of the Association for Information Science and Technology* 和《软件学报》组合的标准化分数超过了 2。此外,在 2019 年, *Journal of Informetrics* 和 *Science* 的组合已经很常见了,具有很强的传统性,而



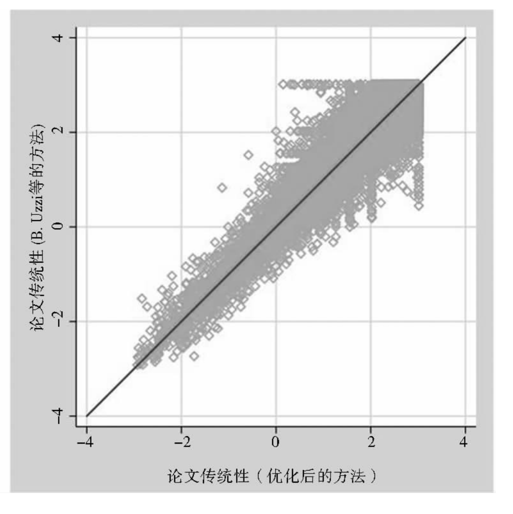


图 7 两种不同方法测得的新颖性和传统性的比较(图中黑色对角线为 y = x)

《情报资料工作》与 *BMJ*(*British Medical Journal*)的组合也比较常见,具有较强的传统性。最后,而相较而言,情报学与人文社会科学的结合具有新颖性,在 2000 年,《情报学报》与《北京大学学报(文科版)》的组合具有一定的新颖性,在 2019 年《情报学报》与《中国社会科学》的组合具有很强的新颖性。而图书馆学与经济学期刊的组合更具有新颖性:在 2000 年,最为罕见的期刊组合之一是《中国图书馆学报》与《经济研究》;而在 2019 年,最为罕见的期刊组合之一是《中国图书馆学报》与 *American Economic Review*。

表 5 不同时代的代表性期刊组合的标准化分数

期刊组合(期刊 – 期刊)	标准化分数(2000 年)	期刊组合(期刊 – 期刊)	标准化分数(2019 年)
中国图书馆学报 – 大学图书馆学报	3. 012 2	中国图书馆学报 – 大学图书馆学报	3. 012 2
Journal of Documentation-Journal of Information Science	2. 921 5	*Information Processing & Management-Scientometrics*	3. 013 4
Journal of American Society for Information Science – 中国图书馆学报	2. 542 6	*Journal of Informetrics-Science*	3. 010 5
情报资料工作 – 知识产权	1. 608 0	*Journal of the Association for Information Science and Technology* – 软件学报	2. 303 5
情报理论与实践 – *College and Research Libraries*	1. 557 0	图书馆 – 知识产权	1. 928 7
图书情报知识 – *Artificial Intelligence*	0. 793 3	情报资料工作 – *BMJ*(*British Medical Journal*)	1. 608 9
图书馆论坛 – 求是	0. 096 3	情报理论与实践 – *Harvard Business Review*	1. 141 1
情报学报 – 人工智能学报	0. 000 0	图书馆建设 – *Information Processing & Management*	0. 786 6
档案学通讯 – 考古	0. 000 0	图书馆论坛 – *Politics*	0. 168 5
图书馆建设 – 情报理论实践	–0. 487 4	档案学通讯 – *PLoS One*	–0. 346 9
情报学报 – 北京大学学报(文科版)	–0. 917 0	图书情报工作 – 中国出版	–1. 002 5
图书情报工作 – *Science*	–1. 434 4	图书情报知识 – *Psychological Review*	–1. 688 5
图书情报知识 – 教育研究	–2. 111 6	大学图书馆学报 – *Science*	–2. 115 4
大学图书馆学报 – 法学研究	–2. 570 4	情报学报 – 中国社会科学	–2. 952 0
中国图书馆学报 – 经济研究	–2. 827 8	中国图书馆学报 – *American Economic Review*	–2. 988 9

在以往的研究中,对新颖性的测量方法的讨论和批判集中在新颖性是否是跨学科程度的一个测量[50]。尽管从表 5 中能大致看出新颖性和跨学科的关系,但为了更好地说明二者的差异,图 8 展示了 2000 年、2010 年和 2019 年 3 个固定年份中图情领域期刊组合(n2000 = 200,n2010 = 295,n2019 = 350)的标准化分数均值和非纯图情领域期刊组合(n2000 = 9 177,n2010 = 47 472,n2009 = 132 352)的标准化分数均值。结果显示,同领域的期刊组合的确会带来很高的标准化分数,即很高的传统性,而不同领域的期刊组合未必会带来高的新颖性。即使是标准化分数均值最低的 2019 年,非纯图情领域期刊组合的标准化分数均值依然大于 0,这说明与随机概率相比,这种组合并不具备新颖性。这些发现与 B. Uzzi 等的研究结果基本一致[9],充分说明优化后的方法也适合用于我国图情领域论文的新颖性和传统性的测量。

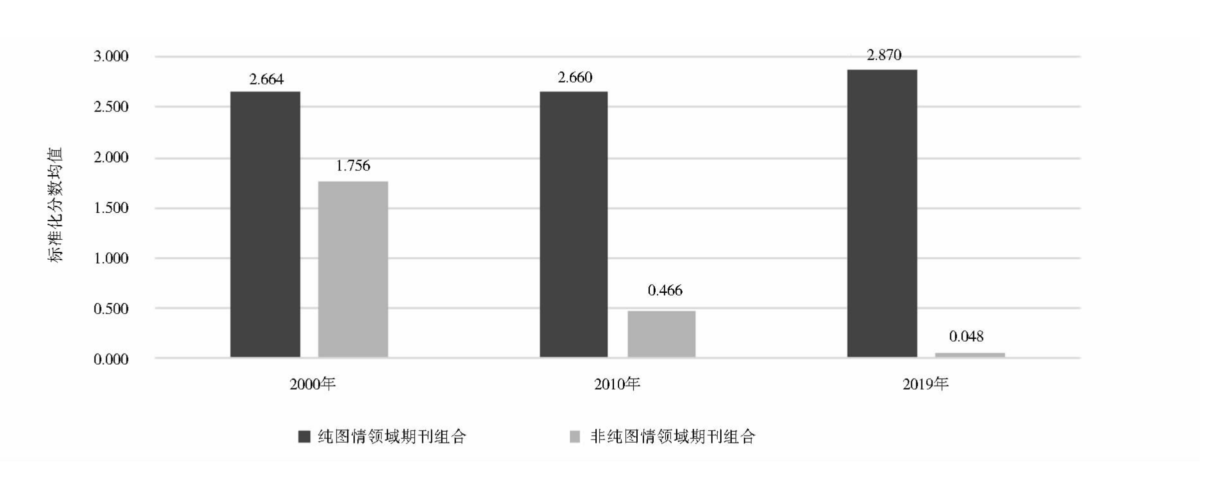


图 8　3 个示例年份的纯图情领域期刊组合和非纯图情领域期刊组合的标准化分数的均值

为了进一步说明跨学科性与新颖性(或传统性)的差异,参照 CSSCI 的专业领域分类计算了不同年份不同领域的期刊与图情领域期刊的组合的标准化分数均值及其 95% 置信区间,结果见图 9。从图 9 可以看出,图情领域的期刊与其他领域的期刊的组合的新颖性存在较大差异,而且随时间变化的趋势明显。相较而言,新闻传播学领域的期刊与图情领域的期刊组合是最为"常见"的跨界组合,标准化分数均值几乎为 0。而文史哲和艺术学领域的期刊与图情领域的期刊组合的标准化分数均值较低;经济学、社会学、心理学、自然资源与环境、人文经济地理等领域的期刊与图情领域的期刊组合标准化分数均值更低。此外,从不同年份的情况来看,这些领域的期刊组合的标准化分数均值有较为明显的差异,这进一步说明新颖性与跨学科性是有差异的。

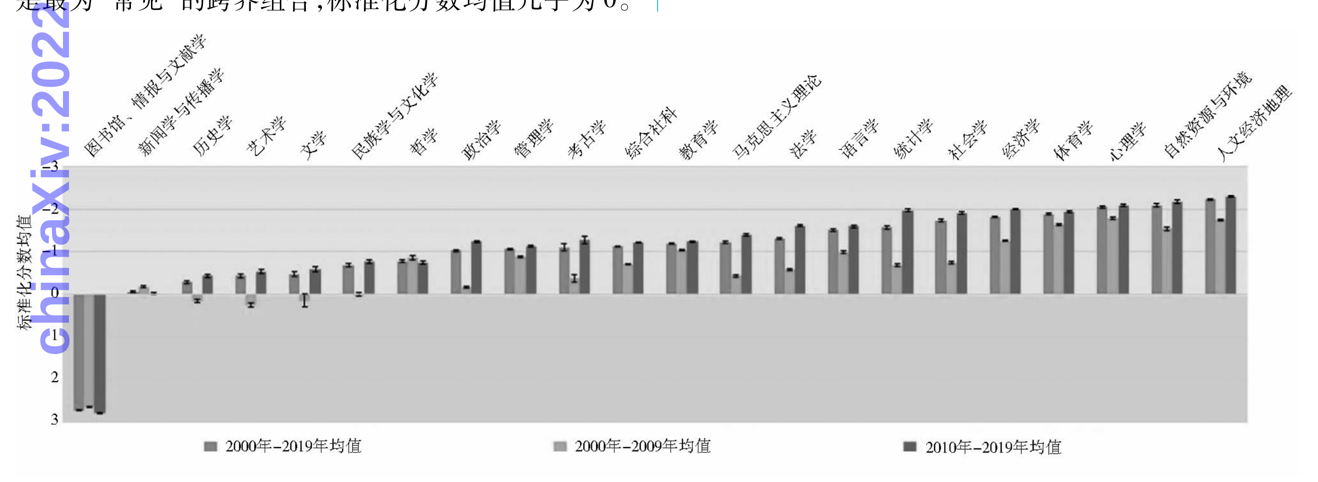


图 9　不同年代图情领域期刊与各领域期刊组合的标准化分数的均值及其 95% 置信区间

注:①CSSCI 的专业类目并不严格按照学科设置。在使用 CSSCI 分类时,参考《学位授予和人才培养学科目录》和《中国图书馆分类法》将原有的宗教学和哲学合并、中国文学和外国文学合并;②标准化分数值均值越小,说明新颖性越强

5.2　我国图情领域的"守正创新"

研究结果验证了假设 1 和假设 2,进而说明,在图情领域的研究中,保持传统性的"守正"、追求新颖性的"创新"都可以成就有影响力的研究。并且,这与库恩提出的科学发展的"范式"有细微区别。在库恩的论述中,学科范式的变化是"割裂的"、剧烈的变化,而图情领域的创新和发展是在传承中进行的。根据回归分析结果(见表 1)可以算出,在 2000 年至 2019 年间,其他因素不变时,论文传统性提高 1 个单位,其成为高被引论文的优势比会提高 33.86%($p<0.0001$);同样地,其他因素不变时,论文的新颖性提高 1 个单位,其成为高被引论文的优势比会提高 11.48%($p<0.0001$)。值得说明的是,对该结果的解读务必非常谨慎。尽管从系数本身来看,传统性的系数大得多,但传

统性的均值为 2.632，而最大值为 3.015。因此，尽管其系数较大，但传统性提升一个单位是不太可能的。而新颖性的情况完全相反，其均值仅为 1.369 277，而最大值同为 3.015，因而提升一个单位是有可能的。所以，结合取值范围和系数来看，新颖性和传统性的影响其实差别不大。此外，这两个变量提升 1 个单位，实际上都是非常大的提升了，以新颖性为例，新颖性为 -1.5排在前 8.18%，而新颖性为 -0.5 却排在前 22.39%。尽管如此，这些结果说明，图情领域的高被引论文都兼具传统性和新颖性，即保持相当的传统性的基础上才能更好地进行创新，进而充分体现了我国图情领域"守正创新"的特点。结合图 5 的边际效应分析，可以进一步了解图情领域论文"守正创新"的特点。显然，新颖性高同时传统性高的论文最有潜力成为高被引的论文，比新颖性低且传统性低的论文成为高被引论文的概率高出近 1.25 倍，比新颖性高传统性低的论文和新颖性低传统性高的论文成为高被引论文的概率也分别提高 44.49% 和 53.91%。这进一步说明在图情领域中，在充分继承学术传统的基础上开展的创新，即守正创新，更具有潜在的学科影响力[51-52]。

对比表 1 中的其他两个模型可以看出，图情领域新颖性的作用在减弱，而传统性的作用在加强。从某种意义上说，这可能是由近年来信息技术的迅速发展和快速迭代而导致的。图情领域的诸多研究高度关注信息技术的新发展，然而在研究新技术的同时，还没能将新技术与图情领域的核心问题紧密融合，形成新的核心研究问题。尽管如此，这种趋势还是说明图情领域的高被引论文通常是在充分利用本学科已有知识之上的传承和发展，并且，注重学科传承的趋势逐渐加强。从另一个角度来看，这也说明图情领域对学术传承的"坚守"的重视，即需要先"守正"才能后"创新"，并且这种"守正"并非因循守旧，而是以寻求创新为导向的[7,19,31]。

研究结果对如何在图情领域"守正创新"具有启示作用。从"守正"来看，一方面，图情领域对学术传承的重视启示我们应当进一步总结、回顾和反思已有的理论和方法。对已有研究成果不断提炼、发展和完善才能形成一般性的、综合的、开创性的新理论、新方法。例如，刘国钧先生在论述建立统一的、符合我国发展需要的图书分类法时，首先系统地梳理了我国图书分类法的发展，然后深度剖析了众所周知的史志目录、杜威法和当时正在使用的一系列分类法的基本情况和不足之处，进而说明新的分类法应当解决的问题和应当具备的特点，从而"有的放矢"地进行创新，创造性地形成了至今仍具有影响力的分类法[53]。

另一方面，图情领域具有很强的实践性，绝大部分研究都来自实践或基于实践。从实践中提炼理论或方法时，往往会形成带有很强的"局部性"的特征，即与实践的具体情境有所关联，而更为深刻的、更具一般性的理论需要我们进一步挖掘学科内涵，以实现对已有知识体系的完善与融合。例如，图书馆学五要素理论是刘国钧先生提出的、面向传统环境中的图书馆的重要理论。在图书馆面临"数智"环境的今天，该理论尽管仍具有指导作用，但也需要与时俱进，既在理论上有所发展，也在实证研究中得到检验，从而使该理论不断完善，形成图书馆学的重要基础理论。

从"创新"来看，图情领域的创新必须紧密围绕本领域关切的核心议题展开。不同学科或领域的知识方法或者不断发展的信息技术都为图情领域的创新提供了丰沃的土壤，但必须将这些"新"内容，根植于图情的核心议题之中，思考新知识、新技术、新方法等对于图情领域的核心研究议题的价值和贡献，选择更具有切题性的突破口来结合"新"内容，才有可能实现有影响力的创新，从而为图情领域的发展做贡献。换言之，应当从本学科领域发展的需要来学习和吸收其他领域成熟的理论体系和研究方法，进而实现突破和创新。此外，结合图 8 和图 9 可以发现，未来图情领域的"守正创新"，可以更多地考虑与一些以往关注较少的领域进行碰撞和融合，如经济学、社会学、心理学、自然资源与环境、人文经济地理等，从个体、群体、社会、外部自然和人文环境等视角进一步探索图情领域关注的现象，发展和完善图情领域已有的规律。

从图 6 中可以发现，在图情领域，团队合作不仅有助于提高研究成果的新颖性，还有助于将研究成果的传统性保持在较高水平。换言之，从学科创新的微观机制而言，团队作者在图情领域"守正创新"的过程中起着重要作用，并且其重要性随时间变迁在不断加强，团队作者对新颖性的影响越来越强于双人作者的影响。这充分说明，依靠研究团队更容易实现图情领域的守正创新[38]。然而，与十几年前的情况类似[36]，目前图情领域论文合作依然是以 2-3 人的小团队合作为主（占到了作者人数规模的 43.96%），3 人及以上的仅占 6.24%。这说明，一方面，图情领域的研究者还未充分认识到或尚未充分挖掘研究团队在学科守正创新中的作用。另一方面，这也可能与图情领域的学科制度有所关联。目前，图情领域的院校在招聘或职称

评定环节认定申请者的研究成果时,往往要求申请者所提交的论文是其以第一作者或通讯作者完成的。这为多人团队合作带来了制度障碍,如何突破这种规则的桎梏,释放研究团队在学科创新中的潜力,也是图情领域未来需要研究解决的重要问题之一。

6 结语

本文通过一种优化后的基于 MCMC 的方法对我国 2000 年至 2019 年 20 年间在中文社会科学引文索引(CSSCI)中收录的图书馆学情报学领域的 70 207 篇研究论文的新颖性、传统性进行了测量;通过回归分析,验证了论文新颖性和传统性对本领域论文学科影响力的作用,分析结果具有很好的统计稳健性。这些研究结果充分说明了我国图情领域守正创新的特点,并对如何在图情领域守正创新具有启示作用,也为理解我国图情领域学术创新规律提供了有力的实证支持。

尽管本文的重点在于探索图情领域的守正创新的规律,但在研究方法上也有所贡献。本文是国内图情领域中最先使用 MCMC 方法的研究。从统计上看,MCMC 方法与以往信息计量研究中使用的方法有着本质区别。以往信息计量研究中使用的方法属于统计中的频率学派,其基本思路是:通过大量的经验研究,得到关于引文分布的参数的统计特征,进而假定其总体分析,然后进行统计推断。例如,引文幂律分布是信息计量中的基本定理。因此,不少研究都希望通过发现现实引文网络的幂律分布特性来检验这一定理。从整体来看,引文幂律分布得到了不同程度的验证[42－43,54],基本上呈现出不同领域引文幂律分布具体参数有所区别,但存在共性[42],同时也存在一些幂律分布不完全成立的情况[54]。即使假定幂律分布是恒成立的,在研究实际中,能使用的数据往往是在不定领域、特定主题或特定年份等一系列条件下收集的。这些限制条件下得到的数据,难以满足大数定律,因而从数据得到的引文分布未必都服从或近似服从幂律分布,或者这些数据来自多个领域,进而出现多个幂律分布的联合分布的情况。无论哪种情况,都会导致现实研究中出现复杂统计分布统计推断的问题。此时的似然函数过于复杂,难以进行最大似然估计。而面对这些问题应当如何开展有效的信息计量分析,还缺乏相对系统深入的研究。

本文采用的 MCMC 是一种基于贝叶斯学派的统计方法,是可以有效解决上述问题的一种方法。与频率学派不同,贝叶斯学派的方法无需假定总体分布的似然函数,而是利用贝叶斯的统计推断方法(如 MCMC),以经验分布为后验分布来解决一些最大似然估计难以操作的问题。例如,在本文中,无需事先知晓 2000－2019 年 CSSCI 收录的我国图情领域研究论文的引文服从何种(幂律)分布及其似然函数的形式,可以通过 MCMC 的方法从观测到的引文关系中构造出与之近似甚至几乎相同的后验分布,并进行相关参数估计,或进行模拟,通过模拟值与观测值的比较来进行统计推断。这种方法更适合引文分布较为复杂、其对应的似然函数不便计算的情况。实际上,贝叶斯统计的方法在样本量相较总体较小或不确定较强的情形下,往往能有更好的统计效力,也能完成一些传统频率学派的方法所不能实现的统计推断。因此,它有效地补充了传统信息计量的方法,并极大地丰富了信息计量的方法。

参考文献:

[1] 吴慰慈,罗志勇.面向 21 世纪图书馆学研究的新趋向[J].中国图书馆学报,2000(6):3－6.
[2] 赖茂生.21 世纪情报学学科的新起点[J].情报学报,2000(1):81.
[3] 赖茂生.信息管理学科发展如何破解瓶颈悖论?[J].大学图书情报学刊,2019,37(5):3－7.
[4] 吴慰慈.对新时代图书馆学研究与学科发展的一点思考——在 2018 年第 12 届全国图书馆学博士生学术论坛上的讲话[J].图书馆杂志,2019,38(6):24.
[5] 张久珍.图情档学科建设需要与新时代发展接轨[J].图书与情报,2020(6):17－18.
[6] 马费成,李志元.新文科背景下我国图书情报学科的发展前景[J].中国图书馆学报, 2020,46(6):4－15.
[7] 孙建军,李阳,裴雷."数智"赋能时代图情档变革之思考[J].图书情报知识,2020(3):22－27.
[8] 柯平.新图情档——新文科建设中的图书情报与档案管理一级学科发展[J].情报资料工作,2021,42(1):15－20.
[9] UZZI B, MUKHERJEE S, STRINGER M, et al. Atypical combinations and scientific impact[J]. Science, 2013,342(6157): 468－472.
[10] WEITZMAN M L. Recombinant growth[J]. Quarterly journal of economics, 1998,113(2):331－360.
[11] MCKNIGHT S L. Unconventional wisdom[J]. Cell, 2009,138(5): 817－819.
[12] SCHILLING M A, GREEN E. Recombinant search and breakthrough idea generation: an analysis of high impact papers in the social sciences[J]. Research policy, 2011,40(10):1321－1331.
[13] ARTS S, FLEMING L. Paradise of novelty-or loss of human capital? exploring new fields and inventive output[J]. Organization

science, 2019,29(6): 1074-1092.
[14] MUKHERJEE S, UZZI B, JONES B, et al. A new method for identifying recombinations of existing knowledge associated with high-impact innovation[J]. Journal of product innovation management, 2016,33(2):224-236.
[15] FOSTER J G, RZHETSKY A, EVANS J A. Tradition and innovation in scientists' research strategies[J]. American sociological review, 2015,80(5):875-908.
[16] WAGNER C S, WHETSELL T A, MUKHERJEE S. International research collaboration: novelty, conventionality, and atypicality in knowledge recombination[J]. Research policy, 2019,48(5): 1260-1270.
[17] LEAHEY E, BECKMAN C M, STANKO T L. Prominent but less productive: the impact of interdisciplinarity on scientists' research [J]. Administrative science quarterly, 2017,62(1):105-139.
[18] WANG J, VEUGELERS R, STEPHAN P. Bias against novelty in science: a cautionary tale for users of bibliometric indicators[J]. Research policy, 2017,46(8): 1416-1436.
[19] 刘兹恒. 再论图书馆学本土化[J]. 图书与情报,2005(4):13-16.
[20] 王世伟. 论智慧图书馆的三大特点[J]. 中国图书馆学报,2012, 38(6):22-28.
[21] 张晓林. 走向知识服务:寻找新世纪图书情报工作的生长点[J]. 中国图书馆学报,2000(5):30-35.
[22] KAPLAN S, VAKILI K. The double-edged sword of recombination in breakthrough innovation[J]. Strategic management journal, 2015,36(10): 1435-1457.
[23] IARIA A, SCHWARZ C, WALDINGER F. Frontier knowledge and scientific production: evidence from the collapse of international science[J]. Quarterly journal of economics, 2018,133(2): 927-991.
[24] HOFSTRA B, KULKARNI V V, MUNOZ-NAJAR GALVEZ S, et al. The diversity-innovation paradox in science[J]. Proceedings of the National Academy of Sciences, 2020,117(17):9284-9291.
[25] BOUDREAU K J, GUINAN E C, LAKHANI K R, et al. Looking across and looking beyond the knowledge frontier: intellectual distance, novelty, and resource allocation in science[J]. Management science, 2016,62(10):2765-2783.
[26] LU S, BARTOL K M, VENKATARAMANI V, et al. Pitching novel ideas to the boss: the interactive effects of employees' idea enactment and influence tactics on creativity assessment and implementation[J]. Academy of management journal, 2019,62(2): 579-606.
[27] CRISCUOLO P, DAHLANDER L, GROHSJEAN T, et al. Evaluating novelty: the role of panels in the selection of R&D projects [J]. Academy of management journal, 2017, 60(2): 433-460.
[28] COONEY G, GILBERT D T, WILSON T D. The novelty penalty: why do people like talking about new experiences but hearing about old ones? [J]. Psychological science, 2017,28(3):380-394.
[29] DEWEY M. The profession[J]. American library journal, 1876, 1(1): 5-6.
[30] 王重民. 論《四庫全书总目》[J]. 北京大学学报(人文科学), 1964(2):63-78.
[31] 王子舟. 学术创新必先从学术史研究入手[J]. 图书情报工作, 2007(3):5.
[32] LEE Y, WALSH J P, & WANG J. Creativity in scientific teams: unpacking novelty and impact[J]. Research policy, 201, 544 (3): 684-697.
[33] TEODORIDIS F. Understanding team knowledge production: the interrelated roles of technology and expertise[J]. Management science, 2018,64(8): 3625-3648.
[34] SINGH J, FLEMING L. Lone inventors as sources of breakthroughs: myth or reality? [J]. Management science, 2010,56 (1): 41-56.
[35] LEAHEY E. From sole investigator to team scientist: trends in the practice and study of research collaboration[J]. Annual review of sociology, 2016,42(1): 81-100.
[36] 魏瑞斌. 我国图书馆学情报学的科研合作现状研究——以CSSCI 1998-2004年数据为例[J]. 图书情报工作,2006(1): 41-43,48.
[37] 李长玲,魏绪秋,崔斌,等. 2004-2013年我国图书情报学科研合作网络结构特征分析[J]. 情报杂志,2015,34(3):119-124,143.
[38] 周晓英,董伟,朱小梅,等. 图书馆学情报学高影响力论文特征及所反映的学科差异分析[J]. 中国图书馆学报,2012,38(4): 99-109.
[39] BORNMANN L. How to analyze percentile citation impact data meaningfully in bibliometrics: the statistical analysis of distributions, percentile rank classes, and top-cited papers[J]. Journal of the American Society for Information Science and Technology, 2013,64(3): 587-595.
[40] BORNMANN L, LEYDESDORFF L, MUTZ R. The use of percentiles and percentile rank classes in the analysis of bibliometric data: opportunities and limits[J]. Journal of informetrics, 2013,7(1): 158-165.
[41] AKSNES D W, LANGFELDT L, WOUTERS P. Citations, citation indicators, and research quality: an overview of basic concepts and theories[J]. SAGE open, 2019, 9(1):215824401982957.
[42] RADICCHI F, FORTUNATO S, CASTELLANO C. Universality of citation distributions: toward an objective measure of scientific impact[J]. Proceedings of the National Academy of Sciences, 2008, 105(45): 17268-17272.
[43] PRICE D D S. A general theory of bibliometric and other cumulative advantage processes[J]. Journal of the American Society for Information Science, 1976,27(5): 292-306.
[44] BARRAT A, BARTHÉLEMY M, PASTOR-SATORRAS R, et al. The architecture of complex weighted networks[J]. Proceedings of the National Academy of Sciences, 2004,101(11):3747-3752.

[45] SMOLINSKY L, LERCHER A, MCDANIEL A. Testing theories of preferential attachment in random networks of citations[J]. Journal of the Association for Information Science and Technology, 2015, 66(10): 2132-2145.
[46] WANG M, YU G, YU D. Measuring the preferential attachment mechanism in citation networks[J]. Physica a, 2008, 387(18): 4692-4698.
[47] TEH Y W, JORDAN M I, BEAL M J, et al. Hierarchical dirichlet processes[J]. Journal of the American Statistical Association, 2006, 101(476): 1566-1581.
[48] 梁兴堃.图情档研究中的回归分析:基本原理[J].图书情报知识,2021,38(3):154-164.
[49] 梁兴堃.图情档研究中的回归分析:常见问题[J].图书情报知识,2021,38(3):165-176.
[50] FONTANA M, IORI M, MONTOBBIO F, et al. New and atypical combinations: an assessment of novelty and interdisciplinarity[J]. Research policy, 2020, 49(7): 104063.
[51] 叶继元, CHEN C M.坚守与拓展:中美图书馆学情报学教育科学定位的思考[J].中国图书馆学报,2007(2):18-23.
[52] 范并思.图书馆学理论道路的迷茫、艰辛与光荣——中国图书馆学暨《中国图书馆学报》六十年[J].中国图书馆学报,2017,43(1):4-16.
[53] 刘国钧.中国图书分类法的发展——附现在中国图书馆图书分类法情况简述[J].图书馆学通讯,1981(2):46-59.
[54] 田盛慧,苏林伟,赵星.多层次施引源项与总被引的幂律关系实证[J].情报学报,2015,34(10):1024-1030.

Novelty, Conventionality, and Scientific Impact of Papers in Library and Information Science in China: Evidence from Papers in CSSCI (2000-2019)

Liang Xingkun
Department of Information Management, Peking University, Beijing 100871

Abstract: [**Purpose/Significance**] With papers in the field of library and information science in China, this study measures the novelty and conventionality of these papers and explores their effects on the scientific impacts of these papers, in order to reveal the law of scientific innovation in this field. [**Method/Process**] This paper adopted a variation of the Markov Chain Monte Carlo (MCMC) approach to measuring novelty and conventionality of 70 207 papers in library and information science ($n = 70\ 207$) from the Chinese Social Science Citation Index (CSSCI) in two decades, specifically, from 2000 to 2019. With robust logistic regression, this paper examined the impacts of papers' novelty and conventionality on their papers' scientific impacts. [**Result/Conclusion**] The results show that, ceteris paribus, the novelty of a paper increases by 1 unit, the odds ratio of the paper becoming a highly cited paper increases by 11% ($p < 0.000\ 1$), and the conventionality of a paper increases by 1 unit, and the odds ratio of the paper becoming a highly cited paper increases by 33% ($p < 0.000\ 1$). The marginal effect analysis shows that papers with high novelty and conventionality are more likely to be highly cited papers than other types of papers. In addition, over time, the impact of novelty on the probability of a paper being highly cited gradually weakens, while the impact of conventionality gradually increases. Meanwhile, author team size has a significant effect on the novelty of the paper, and such a effect increases over time. These findings highlight the characteristics of scientific innovation in the field of library and information in China, and provide novel empirical evidence to understand the law of scientific innovation in the field of library and information in China. Last but not least, this paper, based on Bayesian statistics, also proposes an alternative method for informetrics.

Keywords: novelty conventionality scientific impact hit paper paper citation MCMC team authors Bayesian statistics